PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-091476
(43)Date of publication of application : 27.03.2002

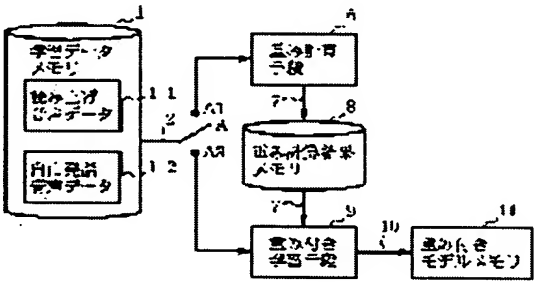
(51)Int.Cl. G10L 15/06
G10L 15/14

(21)Application number : 2000-276944
(22)Date of filing : 12.09.2000

(71)Applicant : MITSUBISHI ELECTRIC CORP
(72)Inventor : HANAZAWA TOSHIYUKI

(54) DEVICE AND METHOD FOR LEARNING VOICE PATTERN MODEL

(57)Abstract:
PROBLEM TO BE SOLVED: To solve the problem in which an equal and robust HMM can not be obtained unless the amounts of learning data of reading voice data 100a and freely uttered voice data 100b are approximately equal for all of the phonemes in the both uttering systems.
SOLUTION: The device is provided with a learning data storage means which stores learning data for different uttering systems, a weight computing means which normalizes the reciprocals of the amounts of data for every system so that the total sum of the reciprocals becomes 1 and computes the normalized reciprocals as the weighting coefficients in accordance with the amount of data for every uttering system in the learning data and a weighted learning means which corrects the differences among the amounts of the data in the uttering systems of the learning data using the weighted coefficients computed by the weight computing means and which obtains the parameters of the voice pattern models corresponding to the learning data.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's

decision of rejection]
[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2002－91476
（P2002－91476A）
(43)公開日　平成14年３月27日(2002.3.27)

(51)Int.Cl.⁷	識別記号	ＦＩ	テーマコード（参考）
Ｇ１０Ｌ　15/06		Ｇ１０Ｌ　3/00	５２１Ｒ　５Ｄ０１５
15/14			５３５Ｚ

審査請求　未請求　請求項の数８　ＯＬ　（全 16 頁）

(21)出願番号　特願2000－276944(P2000－276944)

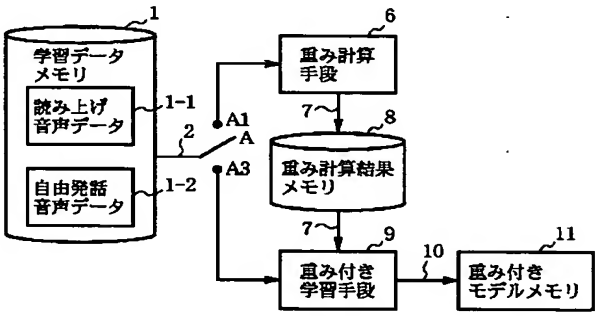
(22)出願日　平成12年９月12日(2000.9.12)

(71)出願人　000006013
三菱電機株式会社
東京都千代田区丸の内二丁目２番３号
(72)発明者　花沢　利行
東京都千代田区丸の内二丁目２番３号　三菱電機株式会社内
(74)代理人　100066474
弁理士　田澤　博昭　（外１名）

Ｆターム(参考)　5D015 GG00 HH23

(54)【発明の名称】　音声パターンモデル学習装置及び音声パターンモデル学習方法

(57)【要約】
【課題】　各音素の全てに対して読み上げ音声データ１００ａと自由発話音声データ１００ｂとの学習データ量が同量程度でなければ、両方の発話様式に対して等しくロバストなＨＭＭを得られないという課題があった。
【解決手段】　異なる発話様式の学習データを格納する学習データ記憶手段と、発話様式ごとの各データ量の逆数に、これらの総和が１となるように正規化したものを、学習データに対する発話様式ごとのデータ量に応じた重み係数として算出する重み計算手段と、この重み計算手段が算出した重み係数を用いて学習データの各発話様式間におけるデータ量の違いを補正しながら、学習データに対応する音声パターンモデルのパラメータを求める重み付き学習手段とを備えた。



【特許請求の範囲】
【請求項１】　異なる発話様式で入力された音声データの音響的特徴を表す複数種類の学習データを格納する学習データ記憶手段と、
上記複数種類の学習データにおける発話様式ごとの各データ量の逆数に、これらの総和が１となるように正規化したものを、上記学習データに対する上記発話様式ごとのデータ量に応じた重み係数として算出する重み計算手段と、
この重み計算手段が算出した上記重み係数を用いて上記学習データの上記各発話様式間におけるデータ量の違いを補正しながら、上記学習データに対応する音声パターンモデルのパラメータを求める重み付き学習手段と
を備えた音声パターンモデル学習装置。
【請求項２】　音声パターンモデルは、隠れマルコフモデルであり、
重み付き学習手段は、重み計算手段が算出した重み係数を学習データから算出した遷移回数期待値に乗じて、上記学習データの各発話様式間におけるデータ量の違いを補正した重み付き遷移回数期待値とし、この重み付き遷移回数期待値を用いて上記隠れマルコフモデルのパラメータを求めることを特徴とする請求項１記載の音声パターンモデル学習装置。
【請求項３】　異なる発話様式で入力された音声データの音響的特徴を表す複数種類の学習データを格納する学習データ記憶手段と、
上記学習データの発話様式ごとに対応する音声パターンモデルのパラメータを求める発話様式別音声パターンモデル学習手段と、
この発話様式別音声パターンモデル学習手段が求めた発話様式ごとのパラメータに対応する発話様式別音声パターンモデルを用いて、上記学習データ記憶手段が格納する上記学習データをクラスタリングし、上記各学習データが属する発話様式のクラスタを決定するクラスタリング手段と、
このクラスタリング手段がクラスタリングした各発話様式のクラスタに属する学習データのデータ量の逆数に、これらの総和が１となるように正規化したものを、上記学習データに対する上記発話様式のクラスタごとのデータ量に応じたクラスタ重み係数として算出するクラスタ重み計算手段と、
このクラスタ重み計算手段が算出した上記クラスタ重み係数を用いて上記学習データの上記各発話様式のクラスタ間におけるデータ量の違いを補正しながら、上記学習データに対応する音声パターンモデルのパラメータを求めるクラスタ重み付き学習手段とを備えた音声パターンモデル学習装置。
【請求項４】　音声パターンモデルは、隠れマルコフモデルであり、
クラスタ重み付き学習手段は、クラスタ重み計算手段が算出したクラスタ重み係数を学習データから算出した遷移回数期待値に乗じて、上記学習データの各発話様式のクラスタ間におけるデータ量の違いを補正したクラスタ重み付き遷移回数期待値とし、このクラスタ重み付き遷移回数期待値を用いて上記隠れマルコフモデルのパラメータを求めることを特徴とする請求項３記載の音声パターンモデル学習装置。
【請求項５】　異なる発話様式で入力された音声データの音響的特徴を表す複数種類の学習データにおける上記発話様式ごとの各データ量の逆数に、これらの総和が１となるように正規化したものを、上記学習データに対する上記発話様式ごとのデータ量に応じた重み係数として算出する重み計算ステップと、この重み計算ステップで算出した重み係数を用いて上記学習データの上記各発話様式間におけるデータ量の違いを補正しながら、上記学習データに対応する音声パターンモデルのパラメータを求める重み付き学習ステップとを備えた音声パターンモデル学習方法。
【請求項６】　音声パターンモデルは、隠れマルコフモデルであり、
重み付き学習ステップにて、重み計算ステップで算出した重み係数を学習データから算出した遷移回数期待値に乗じて、上記学習データの各発話様式間におけるデータ量の違いを補正した重み付き遷移回数期待値とし、この重み付き遷移回数期待値を用いて上記隠れマルコフモデルのパラメータを求めることを特徴とする請求項５記載の音声パターンモデル学習方法。
【請求項７】　異なる発話様式で入力された音声データの音響的特徴を表す複数種類の学習データの上記発話様式ごとに対応する音声パターンモデルのパラメータを求める発話様式別音声パターンモデル学習ステップと、
この発話様式別音声パターンモデル学習ステップにて求めた発話様式ごとのパラメータに対応する発話様式別音声パターンモデルを用いて上記学習データをクラスタリングし、上記各学習データが属する発話様式のクラスタを決定するクラスタリングステップと、
このクラスタリングステップでクラスタリングした各発話様式のクラスタに属する学習データのデータ量の逆数に、これらの総和が１となるように正規化したものを、上記学習データに対する上記発話様式のクラスタごとのデータ量に応じたクラスタ重み係数として算出するクラスタ重み計算ステップと、
このクラスタ重み計算ステップで算出した上記クラスタ重み係数を用いて上記学習データの上記各発話様式のクラスタ間におけるデータ量の違いを補正しながら、上記学習データに対応する音声パターンモデルのパラメータを求めるクラスタ重み付き学習ステップとを備えた音声パターンモデル学習方法。
【請求項８】　音声パターンモデルは、隠れマルコフモデルであり、

クラスタ重み付き学習ステップにて、クラスタ重み計算ステップで算出したクラスタ重み係数を学習データから算出した遷移回数期待値に乗じて、上記学習データの各発話様式のクラスタ間におけるデータ量の違いを補正したクラスタ重み付き遷移回数期待値とし、このクラスタ重み付き遷移回数期待値を用いて上記隠れマルコフモデルのパラメータを求めることを特徴とする請求項７記載の音声パターンモデル学習方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】この発明は朗読調の丁寧な発声の音声や自由発話音声のように発話速度がはやく曖昧な音声などのように発話様式の異なる複数種類の発話の音響特徴を適切にモデル化する音声パターンモデル学習装置及び音声パターンモデル学習方法に関するものである。

【０００２】

【従来の技術】音声認識は、一般に音声データを音響分析して得られる音声データの音響的特徴量である学習データに相当する特徴ベクトルの時系列と、その特徴ベクトルの時系列のパターンをモデル化した音声パターンモデルとのパターンマッチングを行うことにより実現される。この音声パターンモデルとしては、隠れマルコフモデル（Ｈｉｄｄｅｎ　Ｍａｒｋｏｖ　Ｍｏｄｅｌ　以下、ＨＭＭと称する）が用いられることが多い。このＨＭＭでは、大量の学習データを用意することができれば、音声データにおける特徴ベクトルの時系列のパターンを精度よくモデル化することができる。

【０００３】しかしながら、認識対象とする音声データが、学習データとは異なる発話様式で入力されたものであると、認識精度が低下するという問題点があった。特に、連続音声認識においては、テキストを読み上げた音声でＨＭＭを学習しても、対話音声のように自由に発声した音声に対してはロバストなモデルにはならず、充分な認識率が得られないことが知られている。この問題に対処するために、Ｒｉｃｈａｒｄ　Ｐ．　Ｌｉｐｐｍａｎｎ．，ｅｔ　ａｌ．，”Ｍｕｌｔｉ－Ｓｔｙｌｅ　Ｔｒａｉｎｉｎｇ　ｆｏｒ　Ｒｏｂｕｓｔ　Ｉｓｏｌａｔｅｄ－ＷｏｒｄＳｐｅｅｃｈ　Ｒｅｃｏｇｎｉｔｉｏｎ”，ＩＥＥＥ　Ｉｎｔｅｒｎａｔｉｏｎａｌ　Ｃｏｎｆｅｒｅｎｃｅ　ｏｎ　Ａｃｏｕｓｔｉｃｓ，Ｓｐｅｅｃｈ，ａｎｄ　Ｓｉｇｎａｌ　Ｐｒｏｃｅｓｓｉｎｇ，ｖｏｌ．２　ｏｆ　４，ｐｐ．７０５－７０８，１９８７（以下、文献１と称する）では、様々な発話様式で入力された音声データを用いてＨＭＭを学習する方法を提案している。なお、上記文献１では単語音声のモデル化に適用した場合を説明しているが、全く同じ技術を連続音声のモデル化に適用することができる。

【０００４】ここで、連続音声のモデル化に文献１の技術を使用し、テキストを読み上げたような丁寧な音声と、対話音声のように自由に発声した音声の両方に対してロバストなＨＭＭを学習する方法を説明する。本例では学習するＨＭＭは連続分布型のＨＭＭであり、１個のＨＭＭで１個の音素をモデル化するものとする。また、ＨＭＭのトロポジは、例えば図６に示すように４状態のｌｅｆｔ－ｔｏ－ｒｉｇｈｔ型とする。ＨＭＭは、遷移確率a_{ij}と出力確率b_{ij}（Ｘ）とをパラメータとして持つ。ここで、添字ｉｊは、ＨＭＭにおける音声データの音響的特徴の状態が状態ｉからｊに遷移することを示しており、本例では４状態のＨＭＭなので、ｉ＝（１，２，３）、ｊ＝（ｉ，ｉ＋１）となる。また、遷移確率a_{ij}は状態ｉから状態ｊへの遷移が起きる確率であり、出力確率b_{ij}（Ｘ）は状態ｉからｊへの遷移の際に特徴ベクトルＸが出力される確率密度であり、多次元正規分布で表現される。即ち、b_{ij}（Ｘ）を表現するパラメータが平均値m_{ij}と分散v_{ij}であることから、ＨＭＭのパラメータは、遷移確率a_{ij}と出力確率を表現する平均値m_{ij}及び分散v_{ij}となる。このようなＨＭＭのパラメータを求めることをＨＭＭの学習といい、音声データを音響分析して得られる特徴ベクトルの時系列などのＨＭＭの学習に用いるデータを学習データと定義する。

【０００５】図７は文献１に開示された技術を用いた従来の音声パターンモデル学習装置の構成を示すブロック図である。図において、１００は読み上げ音声データ１００ａや自由発話音声データ１００ｂなどのＨＭＭの学習データ１１０を格納する学習データメモリ、１００ａはテキストを読み上げたような丁寧な音声で入力された音声データを音響分析して得られる読み上げ音声データで、１００ｂは対話音声のように自由に発声した音声で入力された音声データを音響分析して得られる自由発話音声データである。１１０は学習データで、読み上げ音声データ１００ａや自由発話音声データ１００ｂのうち、ＨＭＭの学習時に学習手段１２０によって学習データメモリ１００から適宜読み出されてＨＭＭの学習に使用されるデータである。１２０は学習データメモリ１００内に格納される音声データに対するＨＭＭの学習を行う学習手段、１３０は学習手段１２０が学習したＨＭＭのパラメータ、１４０は学習手段１２０が学習したＨＭＭのパラメータ１３０を格納するモデルメモリである。また、学習データメモリ１００には読み上げ音声データ１００ａと自由発話音声データ１００ｂ、及び、学習対象とする音素名の一覧を記した音素テーブルが格納されている。図８は上述した音素テーブルの一例を示す図である。図に示すように、認識対象の音声データを音素ごとに分割し、これらに通し番号を付して管理している。

【０００６】先ず、学習データメモリ１００内の各データの概要について説明する。読み上げ音声データ１００ａは、多様な音素のコンテキストを含んだ文章を多数の話者が読み上げて入力した音声データを音響分析して得

られる特徴ベクトルの時系列と発話内容を示す音素表記から構成される。具体的には、読み上げによって入力された音声データの音声波形を音響分析して得られる特徴ベクトルXの時系列を音素区間ごとに切り出したトークン$X^{(k)}_1$，$X^{(k)}_2$，$X^{(k)}_3$，・・・，$X^{(k)}_{T(k)}$（肩の添字（k）はk番目のトークンであることを意味する）の集合と、各トークンのトークン番号と音素表記を記した読み上げ音声トークンテーブルである。図9は上述した読み上げ音声トークンテーブルを示す図である。図に示すように、各トークンのトークン番号は1番から順番に付すものとする。また、音響分析は、例えばＬＰＣ分析とし、特徴ベクトルXはＬＰＣケプストラムである。さらに、音素区間ごとへの切り出しは、例えば人間がスペクトログラムを観察して行うものとする。

【０００７】自由発話音声データ１００ｂは、多様な話題に対する人対人の対話音声を音響分析して得られる特徴ベクトルXの時系列と発話内容を示す音素表記から構成される。具体的には、自由発話によって入力された音声データの音声波形を音響分析して得られる特徴ベクトルの時系列を音素区間ごとに切り出したトークン$X^{(k)}_1$，$X^{(k)}_2$，$X^{(k)}_3$，・・・，$X^{(k)}_{T(k)}$（肩の添字（k）はk番目のトークンであることを意味する）の集合と、各トークンのトークン番号と音素表記を記した自由発話音声トークンテーブルである。図１０は上述した自由発話音声データトークンテーブルを示す図である。図に示すように、自由発話音声データ１００ｂ中のトークン番号は、読み上げ音声トークンテーブルの最終番号に続く番号を付与する。音響分析は読み上げ音声データと同様に例えばＬＰＣ分析とし、特徴ベクトルXはＬＰＣケプストラムである。音素区間ごとへの切り出しは、読み上げ音声データ１００ａと同様に、例えば人間がスペクトログラムを観察して行うものとする。読み上げ音声データ１００ａは、テキストを読み上げているので比較的丁寧で明瞭な発声であるのに対し、自由発話音声データ１００ｂは人対人の自然な対話音声であるため音素の特徴ベクトルの変形が激しいのが特徴である。

【０００８】次にHMMを学習する動作について説明する。学習手順１　学習手段１２０は、学習データメモリ１００が保持する音素テーブルを読み込み、この音素テーブルの記述にしたがって、学習対象とする音素を選択する。学習手段１２０は、音素テーブルが、例えば図８のように記述されている場合、先頭の音素である／ａ／を学習対象として選択する。

【０００９】学習手順２　学習手段１２０は、選択した音素と一致する音素表記を持つ全てのトークンの特徴ベクトルの時系列を学習データメモリ１００から読み込む。このとき、読み上げ音声データ１００ａのトークンと自由発話音声データ１００ｂのトークンを両方読み込む。ここで、読み込んだトークンの総数をKとして、読み込んだ各トークン$X^{(k)}_1$，$X^{(k)}_2$，$X^{(k)}_3$，・・・，$X^{(k)}_{T(k)}$に対する遷移回数期待値$\gamma^{(k)}(i,j,t)$を計算する。但し、肩の添字（k）は読み込んだ全てのトークン中でk（k＝1，2，3，・・・K）番目のトークンであることを意味する。Xはトークンを構成する特徴ベクトル、T（k）はk番目のトークンを構成する特徴ベクトルの数とする。また、（i，j，t）はトークンのt番目の特徴ベクトル$X^{(k)}_t$でHMMの状態iから状態jへ遷移したことを示すものとする。この遷移回数期待値$\gamma^{(k)}(i,j,t)$は、例えばフォワード・バックワードアルゴリズムを用いて計算する。

【００１０】次に、遷移回数期待値$\gamma^{(k)}(i,j,t)$を用いて、（１）～（３）式によってHMMのパラメータである遷移確率a_{ij}、平均値m_{ij}及び分散v_{ij}を学習する。HMMは、図６に示したように、４状態なのでi＝（１，２，３）、j＝（i，i＋１）である。

【数１】

$$a_{ij}=\frac{\sum_{k=1}^{K}\sum_{t=1}^{T(k)}\gamma^{(k)}(i,j,t)}{\sum_{k=1}^{K}\sum_{j=i}^{i+1}\sum_{t=1}^{T(k)}\gamma^{(k)}(i,j,t)} \qquad (1)$$

【数２】

$$m_{ij}=\frac{\sum_{k=1}^{K}\sum_{t=1}^{T(k)}\gamma^{(k)}(i,j,t)*X_t^{(k)}}{\sum_{k=1}^{K}\sum_{t=1}^{T(k)}\gamma^{(k)}(i,j,t)} \qquad (2)$$

【数３】

$$v_{ij} = \frac{\sum_{k=1}^{K}\sum_{t=1}^{T^{(k)}}\gamma^{(k)}(i,j,t)*(m_{ij}-X_t^{(k)})^2}{\sum_{k=1}^{K}\sum_{t=1}^{T^{(k)}}\gamma^{(k)}(i,j,t)} \qquad (3)$$

学習対象の音素に対するＨＭＭの学習を終了すると、学習手段１２０は学習を終了したモデルのパラメータ１３０である、遷移確率a_{ij}、平均値m_{ij}、分散v_{ij}、及び、その音素表記をモデルメモリ１４０に送出する。モデルメモリ１４０は、学習を終了した上記モデルのパラメータ１３０及びその音素表記を保持する。

【００１１】学習手順３　学習手段１２０は、学習データメモリ１００が保持する音素テーブルに存在する全ての音素に対してモデルの学習が終了するまで、上述した手順で学習対象とする音素を音素テーブルに記述されている順序にしたがって選択し、上記学習手順２を繰り返す。以上で音素モデルの学習を終了する。

【００１２】このように文献１に示された従来技術では、音響的特徴の異なる種々の発話様式の学習データを用いることにより、種々の発話様式の音声に対してロバストなＨＭＭを得ることを目的としている。本例では読み上げ音声データ１００ａと自由発話音声データ１００ｂとの両方に対してロバストなＨＭＭを得ることを目的としている。

【００１３】

【発明が解決しようとする課題】従来の音声パターンモデル学習装置は以上のように構成されているので、各音素の全てに対して読み上げ音声データ１００ａと自由発話音声データ１００ｂとの学習データ量が同量程度でなければ、両方の発話様式に対して等しくロバストなＨＭＭを得られないという課題があった。

【００１４】上記課題について具体的に説明する。自由発話音声データ１００ｂは、一般的に多様な話題に対する人対人の対話音声を音響分析して得ることからデータ収集が困難であり、読み上げ音声データ１００ａと比較してＨＭＭの学習に使用する学習データ１１０のデータ量が少なくなる。また、読み上げ音声データ１００ａは、発話内容を完全に指定して音声を収録することができるが、自由発話音声データ１００ｂは、一般的に人対人の自由発話音声データ等を収録するので発話内容を完全に指定することができない。従って、読み上げ音声とは音素の出現頻度が異なったものになる。ＨＭＭの学習は、（１）～（３）式に示すように最尤推定に基づいていることから、データ量と音素の出現頻度が異なる学習データを両方用いて、文献１に示された従来技術によってＨＭＭを学習すると、読み上げ音声の特徴と自由発話音声データとの音響的特徴が均等にモデル化されないという不具合があった。即ち、読み上げ音声データ１００ａの方が学習データ量が多くなり、ＨＭＭの学習結果が読み上げ音声データ１００ａの音響的特徴に近いモデルとなり、自由発話音声データ１００ｂの音響的特徴はモデル化されにくくなる。

【００１５】一方、上記課題を解決する従来の技術として、特開平４－３２６４００号公報（以下、文献２と称する）に開示される音響モデル構成方法がある。この文献２に開示される技術は、読み上げ音声データと自由発話音声データとで、別々にＨＭＭを学習する（文献１による技術では、上述したように学習手段１２０が読み上げ音声データ１００ａと自由発話音声データ１００ｂとを両方読み込んで、両者を考慮してＨＭＭの学習を行っている）。このあと、音声認識時に各ＨＭＭの尤度を計算して加重平均を取る。この加重平均により算出された尤度を最適な尤度として設定し、この尤度に対応するパラメータによるＨＭＭを、音声データに最適化された音声パターンモデルとして決定する。しかしながら、上記文献２による技術を使用した音声パターン学習装置では、パラメータを求めるＨＭＭの数が読み上げ音声データと自由発話音声データとで２倍になり、さらに、ＨＭＭパラメータの記憶領域とＨＭＭの尤度計算の演算量も２倍になるという課題があった。これにより、上記演算量と記憶領域とを提供できる程のハードウェア資源を要することからコスト的にも不利であった。

【００１６】この発明は上記のような課題を解決するためになされたもので、読み上げ音声と自由発話音声との学習データ量が異なる場合でも両方の発話様式に対してＨＭＭの数を増加させることなく、ロバストなＨＭＭを求めることができる音声パターンモデル学習装置を得ることを目的とする。

【００１７】

【課題を解決するための手段】この発明に係る音声パターンモデル学習装置は、異なる発話様式で入力された音声データの音響的特徴を表す複数種類の学習データを格納する学習データ記憶手段と、複数種類の学習データにおける発話様式ごとの各データ量の逆数に、これらの総和が１となるように正規化したものを、学習データに対する発話様式ごとのデータ量に応じた重み係数として算出する重み計算手段と、この重み計算手段が算出した重み係数を用いて学習データの各発話様式間におけるデータ量の違いを補正しながら、学習データに対応する音声パターンモデルのパラメータを求める重み付き学習手段とを備えるものである。

【００１８】この発明に係る音声パターンモデル学習装置は、音声パターンモデルが隠れマルコフモデルであり、重み付き学習手段は、重み計算手段が算出した重み係数を学習データから算出した遷移回数期待値に乗じ

て、学習データの各発話様式間におけるデータ量の違いを補正した重み付き遷移回数期待値とし、この重み付き遷移回数期待値を用いて隠れマルコフモデルのパラメータを求めるものである。

【００１９】この発明に係る音声パターンモデル学習装置は、異なる発話様式で入力された音声データの音響的特徴を表す複数種類の学習データを格納する学習データ記憶手段と、学習データの発話様式ごとに対応する音声パターンモデルのパラメータを求める発話様式別音声パターンモデル学習手段と、この発話様式別音声パターンモデル学習手段が求めた発話様式ごとの音声パターンモデルのパラメータに対応する発話様式別音声パターンモデルを用いて、学習データ記憶手段が格納する学習データをクラスタリングし、各学習データが属する発話様式のクラスタを決定するクラスタリング手段と、このクラスタリング手段がクラスタリングした各発話様式のクラスタに属する学習データのデータ量の逆数に、これらの総和が１となるように正規化したものを、学習データに対する発話様式のクラスタごとのデータ量に応じたクラスタ重み係数として算出するクラスタ重み計算手段と、このクラスタ重み計算手段が算出したクラスタ重み係数を用いて学習データの各発話様式のクラスタ間におけるデータ量の違いを補正しながら、学習データに対応する音声パターンモデルのパラメータを求めるクラスタ重み付き学習手段とを備えるものである。

【００２０】この発明に係る音声パターンモデル学習装置は、音声パターンモデルが隠れマルコフモデルであり、クラスタ重み付き学習手段は、クラスタ重み計算手段が算出したクラスタ重み係数を学習データから算出した遷移回数期待値に乗じて、学習データの各発話様式のクラスタ間におけるデータ量の違いを補正したクラスタ重み付き遷移回数期待値とし、このクラスタ重み付き遷移回数期待値を用いて隠れマルコフモデルのパラメータを求めるものである。

【００２１】この発明に係る音声パターンモデル学習方法は、異なる発話様式で入力された音声データの音響的特徴を表す複数種類の学習データにおける発話様式ごとの各データ量の逆数にこれらの総和が１となるように正規化したものを、学習データに対する発話様式ごとのデータ量に応じた重み係数として算出する重み計算ステップと、この重み計算ステップで算出した重み係数を用いて学習データの各発話様式間におけるデータ量の違いを補正しながら、学習データに対応する音声パターンモデルのパラメータを求める重み付き学習ステップとを備えるものである。

【００２２】この発明に係る音声パターンモデル学習方法は、音声パターンモデルが隠れマルコフモデルであり、重み付き学習ステップにて、重み計算ステップで算出した重み係数を学習データから算出した遷移回数期待値に乗じて、学習データの各発話様式間におけるデータ量の違いを補正した重み付き遷移回数期待値とし、この重み付き遷移回数期待値を用いて隠れマルコフモデルのパラメータを求めるものである。

【００２３】この発明に係る音声パターンモデル学習方法は、異なる発話様式で入力された音声データの音響的特徴を表す複数種類の学習データの発話様式ごとに対応する音声パターンモデルのパラメータを求める発話様式別音声パターンモデル学習ステップと、この発話様式別音声パターンモデル学習ステップにて求めた発話様式ごとの音声パターンモデルのパラメータに対応する発話様式別音声パターンモデルを用いて学習データをクラスタリングして、各学習データが属する発話様式のクラスタを決定するクラスタリングステップと、このクラスタリングステップでクラスタリングした各発話様式のクラスタに属する学習データのデータ量の逆数に、これらの総和が１となるように正規化したものを、学習データに対する発話様式のクラスタごとのデータ量に応じたクラスタ重み係数として算出するクラスタ重み計算ステップと、このクラスタ重み計算ステップで算出したクラスタ重み係数を用いて学習データの各発話様式のクラスタ間におけるデータ量の違いを補正しながら、学習データに対応する音声パターンモデルのパラメータを求めるクラスタ重み付き学習ステップとを備えるものである。

【００２４】この発明に係る音声パターンモデル学習方法は、音声パターンモデルが隠れマルコフモデルであり、クラスタ重み付き学習ステップにて、クラスタ重み計算ステップで算出したクラスタ重み係数を学習データから算出した遷移回数期待値に乗じて、学習データの各発話様式のクラスタ間におけるデータ量の違いを補正したクラスタ重み付き遷移回数期待値とし、このクラスタ重み付き遷移回数期待値を用いて隠れマルコフモデルのパラメータを求めるものである。

【００２５】

【発明の実施の形態】以下、この発明の実施の一形態を説明する。

実施の形態１．図１はこの発明の実施の形態１による音声パターンモデル学習装置の構成を示すブロック図である。図において、１は読み上げ音声データ１－１や自由発話音声データ１－２などのＨＭＭの学習データ２を格納する学習データメモリ（学習データ記憶手段）、１－１はテキストを読み上げたような丁寧な音声で入力された音声データを音響分析して得られる読み上げ音声データ（学習データ）で、１－２は対話音声のように自由に発声した音声で入力された音声データを音響分析して得られる自由発話音声データ（学習データ）である。具体的に読み上げ音声データ１－１及び自由発話音声データ１－２を説明すると、読み上げ音声データ１－１は、読み上げ音声によって入力された音声データの音声波形を音響分析して得られる特徴ベクトルＸの時系列を音素区間ごとに切り出したトークンの集合と、各トークンのト

ークン番号と音素表記を記した読み上げ音声トークンテーブルを指している。また、自由発話音声データ１－２は、自由発話によって入力された音声データの音声波形を音響分析して得られる特徴ベクトルの時系列を音素区間ごとに切り出したトークンの集合と、各トークンのトークン番号と音素表記を記した自由発話音声トークンテーブルを指している。２は学習データで、読み上げ音声データ１－１や自由発話音声データ１－２のうち重み計算手段６や重み付き学習手段９によって学習データメモリ１から適宜読み出されてＨＭＭの学習に使用されるデータである。

【００２６】６は読み上げ音声データ１－１や自由発話音声データ１－２から読み出した学習データ２の各データ量から学習データ２に対する発話様式（読み上げ音声、自由発話音声）ごとのデータ量に応じた重み係数を算出する重み計算手段で、７は重み計算手段６が算出した重み係数である。８は重み計算手段６が算出した重み係数７を格納する重み計算結果メモリ、９は重み計算手段６が算出した重み係数７を用いて、学習データ２の各発話様式（読み上げ音声、自由発話音声）間におけるデータ量の違いを補正しながら、学習データ２に対する音声パターンモデルのパラメータを求める重み付き学習手段、１０は重み付き学習手段９によって学習したＨＭＭのパラメータ（学習データ２に対応する音声パターンモデルのパラメータ）を含むＨＭＭパラメータ情報、１１は重み付き学習手段９が学習データ２の各発話様式（読み上げ音声、自由発話音声）間におけるデータ量の違いを補正しながら求めたＨＭＭパラメータ情報１０を格納する重み付きモデルメモリである。なお、この実施の形態１で学習するＨＭＭは、上述した従来技術と同様に連続分布型のＨＭＭとし、１個のＨＭＭで１個の音素をモデル化するものとする。ＨＭＭのトポロジーは図６に示すような４状態のｌｅｆｔ－ｔｏ－ｒｉｇｈｔ型とする。

【００２７】次に動作について説明する。この実施の形態１による音声パターンモデル学習装置の学習手順は、（１）重み係数の計算（重み計算ステップ）、（２）重み付きモデルの学習（重み付き学習ステップ）の２段階に分かれる。先ず、重み係数の計算手順について説明する。重み係数の計算を行う場合、学習データメモリ１の出力端子Ａを重み計算手段６の入力端子Ａ１に接続する。この接続状態で重み係数の計算を以下の手順で行う。

【００２８】重み係数の計算手順１　重み計算手段６は、上記接続経路を介して学習データメモリ１が保持する音素テーブルを読み込み、この音素テーブルに記述の音素に付された順番に従って、重み係数の計算対象とする音素を選択する。ここで、音素テーブルが、例えば図８のように記述されている場合、通し番号が先頭の音素である／ａ／を重み係数の計算対象として選択する。

【００２９】重み係数の計算手順２　次に、重み計算手段６は、学習データメモリ１が保持する読み上げ音声トークンテーブル（学習データ２）と自由発話音声トークンテーブル（学習データ２）とに記載されているトークンのうち、重み係数の計算対象音素として選択した音素表記を持つトークンの数を上記両トークンテーブルで別々に数え上げる。ここで、読み上げ音声トークンテーブルに対するトークン（学習データ２）の数え上げ数をＣＲ、自由発話音声トークンテーブルに対するトークン（学習データ２）の数え上げ数をＣＳとすると、読み上げ音声データ１－１に対する重み係数ＷＲと自由発話音声データデータ１－２に対する重み係数ＷＳとは、下記（４）、（５）式にしたがって計算される。これら（４）、（５）式から分かるとおり、読み上げ音声データ１－１に対する重み係数ＷＲと自由発話音声データデータ１－２に対する重み係数ＷＳとの比は、読み上げ音声データ１－１に属するトークンの数ＣＲと自由発話音声データデータ１－２に属するトークンの数ＣＳの比に反比例する値となっている。

【数４】

$$WR = \frac{CS}{CR + CS} \qquad (4)$$

【数５】

$$WS = \frac{CR}{CR + CS} \qquad (5)$$

重み計算手段６は、当該音素の音素表記である／ａ／とともに、算出された重み係数ＷＲ、ＷＳを重み計算結果メモリ８に送出する。これにより、重み計算結果メモリ８は音素表記である／ａ／と重み係数ＷＲ、ＷＳとを保持する。

【００３０】重み係数の計算手順３　重み計算手段６は、図８に示した音素テーブルの表記の順番、即ち、／ｉ／～／ｚ／の順番に上述した重み係数の計算手順２を繰り返し、各音素ごとに読み上げ音声データ１－１に対する重み係数ＷＲと自由発話音声データデータ１－２に対する重み係数ＷＳを計算して、当該音素の音素表記とともに重み係数ＷＲ、ＷＳを重み計算結果メモリ８に送出する。重み計算結果メモリ８は音素表記と重み係数ＷＲ、ＷＳとを保持する。図２は実施の形態１による音声パターンモデル学習装置における重み計算結果メモリに保持された重み係数を示す図である。図において、音素

／a／に対する重み係数値は、WR＝0．25、WS＝0．75となっているが、これは音素／a／に対する読み上げ音声データ1－1のトークン数と自由発話音声データ1－2のトークン数の比が0．75：0．25、即ち、読み上げ音声データ1－1のトークン数が自由発話音声データ1－2のトークン数の3倍であることを意味する。このようにして学習データに対する発話様式ごとのデータ量に応じた重み係数の算出が完了する。

【0031】次に重み付きモデルの学習手順を説明する。重み付きモデルの学習を開始する前に、学習データメモリ1の出力端子Aを重み付き学習手段9の入力端子A3に接続する。この接続状態で重み付きモデルの学習を以下の手順で行う。

【0032】重み付きモデルの学習手順1　重み付き学習手段9は、上記接続経路を介して学習データメモリ1が保持する音素テーブルを読み込み、この音素テーブルの記述の音素に付された順番に従って、学習対象とする音素を選択する。ここで、音素テーブルが、例えば図8のように記述されている場合、通し番号が先頭の音素である／a／を学習対象として選択する。

【0033】重み付きモデルの学習手順2　次に、重み付き学習手段9は、上述のようにして選択した音素と一致する音素表記を持つ全てのトークンの特徴ベクトルの時系列である学習データ2を学習データメモリ1から読み込む。この際、読み上げ音声データ1－1中のトークンと自由発話音声データ1－2中のトークンを両方読み込む。読み込んだトークンの総数をKとする。また、重み付き学習手段9は、重み計算結果メモリ8から上記学習手順1で選択した音素に対する読み上げ音声データのWRと自由発話音声データデータの重み係数WSとを読み込む。

【0034】重み付きモデルの学習手順3　重み付き学習手段9は、読み込んだ各トークン$X_{(k)1}$，$X_{(k)2}$，$X_{(k)3}$，・・・，$X_{(k)T(k)}$に対する遷移回数期待値$\gamma_{(k)}(i,j,t)$を計算する。ここで、肩の添字（k）は読み込んだ全てのトークン中でk（k＝1，2，3，・・・，K）番目のトークンであることを意味する、Xはトークンを構成する特徴ベクトル、T（k）はk番目のトークンを構成する特徴ベクトルの数とする。また、（i，j，t）は、トークンのt番目の特徴ベクトル$X_{(k)t}$を出力して、HMMの状態iから状態jへ遷移したことを示すものとする。この遷移回数期待値$\gamma_{(k)}(i,j,t)$は、従来技術と同様に例えばフォワード・バックワードアルゴリズムを用いて計算することができる。重み付き学習手段9は、上記遷移回数期待値$\gamma_{(k)}(i,j,t)$を用いて重み付き遷移回数期待値$\gamma_{(k)w}(i,j,t)$を（6）式に従って算出する。

【数6】

$$\gamma_W^{(k)}(i,j,t) = \begin{cases} WR * \gamma^{(k)}(i,j,t) & (k\text{番目のトークンが読み上げ音声データ中のトークンの場合}) \\ WS * \gamma^{(k)}(i,j,t) & (k\text{番目のトークンが自由発話音声データ中のトークンの場合}) \end{cases} \quad (6)$$

重み付き学習手段9は、算出した重み付き遷移回数期待値$\gamma_{(k)w}(i,j,t)$を用いて下記（7）～（9）式に従って、HMMのパラメータを学習する。ここで、HMMは、図6に示すような4状態とすることから、i＝（1，2，3）、j＝（i，i＋1）となる。

【数7】

$$a_{ij}^{W} = \frac{\sum_{k=1}^{K}\sum_{t=1}^{T(k)}\gamma_W^{(k)}(i,j,t)}{\sum_{k=1}^{K}\sum_{j=i}^{i+1}\sum_{t=1}^{T(k)}\gamma_W^{(k)}(i,j,t)} \quad (7)$$

【数8】

$$m_{ij}^{W} = \frac{\sum_{k=1}^{K}\sum_{t=1}^{T(k)}\gamma_W^{(k)}(i,j,t) * X_t^{(k)}}{\sum_{k=1}^{K}\sum_{t=1}^{T(k)}\gamma_W^{(k)}(i,j,t)} \quad (8)$$

【数9】

$$v_{ij}^{W} = \frac{\sum_{k=1}^{K}\sum_{t=1}^{T(k)}\gamma_{W}^{(k)}(i,j,t)*(m_{ij}^{W}-X_{t}^{(k)})^{2}}{\sum_{k=1}^{K}\sum_{t=1}^{T(k)}\gamma_{W}^{(k)}(i,j,t)} \qquad (9)$$

【００３５】学習を終了すると、重み付き学習手段９は、遷移確率a_{ij}^{W}、平均値m_{ij}^{W}、分散v_{ij}^{W}及びその音素表記を重み付きモデルメモリ１１に送出する。重み付きモデルメモリ１１は、学習を終了した上記HMMのパラメータとその音素表記とをHMMパラメータ情報１０として保持する。

【００３６】重み付きモデルの学習手順４　重み付き学習手段９は、学習データメモリ１が保持する音素テーブルに存在する全ての音素に対してHMMの学習が終了するまで学習対象とする音素を音素テーブルに記述されている順序にしたがって選択し、上記重み付きモデルの学習手順２～３を繰り返す。

【００３７】ここで、実施の形態１と従来技術との違いについて説明する。（７）～（９）式において、通常の遷移回数期待値$\gamma^{(k)}(i,j,t)$の代わりに重み付き遷移回数期待値$\gamma_{W}^{(k)}(i,j,t)$を用いることである。この重み付き遷移回数期待値$\gamma_{W}^{(k)}(i,j,t)$は（６）式に示したとおり当該トークンが読み上げ音声データ１－１中のトークンの場合は重み係数WRをかけ、自由発話音声データ１－２中のトークンの場合には重み係数WSをかけて計算する。重み係数WRとWSとは、（４）、（５）式に示したとおり読み上げ音声データ１－１と自由発話音声データ１－２のトークン数に反比例する値となっている。従って、重み付き遷移回数期待値$\gamma_{W}^{(k)}(i,j,t)$を用いることによって、読み上げ音声データ１－１に属するトークン数と自由発話音声データ１－２に属するトークン数との不均衡を補正し、両発話様式の音響的特徴を均等に反映したHMMを学習することができる。即ち、読み上げ音声と自由発話音声との両方に対してロバストなHMMを得ることができる。

【００３８】以上のように、この実施の形態１によれば、異なる発話様式（読み上げ音声、自由発話音声）で入力された音声データの音響的特徴を表す読み上げ音声データ１－１、自由発話音声データ１－２から抽出した学習データ２の各データ量の逆数１／CR、１／CSに、これらの総和が１となるように正規化したもの（各逆数１／CR、１／CSにCR・CS／（CR+CS）を乗算する、（４）、（５）式参照）を、学習データ２に対する発話様式ごとのデータ量に応じた重み係数WR、WSとして算出し、この重み係数WR、WSを用いて学習データ２の各発話様式間におけるデータ量の違いを補正しながら、学習データ２に対応する音声パターンモデルのパラメータを求めるので、読み上げ音声データ１－１に属するトークン数と自由発話音声データ１－２に属するトークン数との不均衡を補正し、両発話様式の音響的特徴を均等に反映した音声パターンモデルを学習することができ、読み上げ音声と自由発話音声との両方に対してロバストな音声パターンモデルを得ることができるという効果が得られる。

【００３９】また、文献２による従来技術と異なり、読み上げ音声と自由発話音声とで別々に音声パターンモデルの学習をすることがないので、音声パターンモデルの数を増加させない。これにより、文献２による従来技術と比較して高性能なハードウェア資源を要することがなく、コスト的にも有利である。

【００４０】また、この実施の形態１によれば、音声パターンモデルが隠れマルコフモデルであり、重み係数WR、WSを学習データ２から算出した遷移回数期待値$\gamma^{(k)}(i,j,t)$に乗じて、学習データ２の各発話様式間におけるデータ量の違いを補正した重み付き遷移回数期待値$\gamma_{W}^{(k)}(i,j,t)$とし、この重み付き遷移回数期待値$\gamma_{W}^{(k)}(i,j,t)$を用いて、隠れマルコフモデルのパラメータ遷移確率a_{ij}^{W}、平均値m_{ij}^{W}、分散v_{ij}^{W}を求めるので、読み上げ音声データ１－１に属するトークン数と自由発話音声データ１－２に属するトークン数との不均衡を補正し、両発話様式の音響的特徴を均等に反映したHMMを学習することができ、読み上げ音声と自由発話音声との両方に対してロバストなHMMを得ることができるという効果が得られる。

【００４１】なお、上記実施の形態１では、HMMで音素をモデル化する場合を説明したが、音節など他の音響単位でも構わない。

【００４２】実施の形態２．この実施の形態２は、異なる発話様式で入力された音声データの音響的特徴を表す複数種類の学習データの発話様式ごとに対応する音声パターンモデルのパラメータを求めて、これら発話様式ごとの音声パターンモデルのパラメータに対応する発話様式別音声パターンモデルを用いて学習データをクラスタリングして、各学習データが属する発話様式のクラスタを決定し、クラスタリングした各発話様式のクラスタに属する学習データのデータ量の逆数に、これらの総和が１となるように正規化したものを、学習データに対する発話様式のクラスタごとのデータ量に応じたクラスタ重み係数として算出し、このクラスタ重み係数を用いて学習データの各発話様式のクラスタ間におけるデータ量の違いを補正しながら、学習データに対応する音声パターンモデルのパラメータを求めるものである。

【００４３】図３はこの発明の実施の形態２による音声

パターンモデル学習装置の構成を示すブロック図である。図において、３は読み上げ音声データ１－１及び自由発話音声データ１－２からの学習データ２の発話様式ごとに対応するＨＭＭのパラメータを求める学習手段（発話様式別音声パターンモデル学習手段）である。１２は学習手段３が読み上げ音声データ１－１を用いて学習したＨＭＭ（発話様式別音声パターンモデル）のパラメータを含むＨＭＭパラメータ情報で、１３は学習手段３からのＨＭＭパラメータ情報１２に対応するＨＭＭを格納する読み上げ音声モデルメモリである。１４は学習手段３が自由発話音声データ１－２を用いて学習したＨＭＭ（発話様式別音声パターンモデル）のパラメータを含むＨＭＭパラメータ情報で、１５は学習手段３からのＨＭＭパラメータ情報１４に対応するＨＭＭを格納する自由発話音声モデルメモリである。１６はクラスタリング手段で、学習手段３が発話様式別（読み上げ音声データ１－１、自由発話音声データ１－２）に求めたＨＭＭを用いて学習データメモリ１が格納する学習データ２をクラスタリングして、各学習データ２が属する発話様式のクラスタを決定する。１７はクラスタリング手段１６が決定した学習データ２が属する発話様式のクラスタを各学習データ２ごとに対応付けたデータであるクラスタリング結果、１８はクラスタリング結果１７を格納するクラスタリング結果メモリである。

【００４４】１９はクラスタリング結果メモリ１８が格納するクラスタリング結果１７から各発話様式のクラスタに属する学習データ２のデータ量を計算し、学習データ２に対する発話様式のクラスタごとのデータ量に応じたクラスタ重み係数を算出するクラスタ重み計算手段、２０はクラスタ重み計算手段１９が算出した学習データ２のクラスタ重み係数を各学習データ２ごとに対応付けたデータであるクラスタ重み計算結果、２１はクラスタ重み計算手段１９からのクラスタ重み計算結果２０を格納するクラスタ重み計算結果メモリである。２２はクラスタ重み付き学習手段であって、クラスタ重み計算結果メモリ２１から読み出したクラスタ重み計算結果２０とクラスタリング結果メモリ１８から読み出したクラスタリング結果１７とを用いて学習データ２の各発話様式のクラスタ間におけるデータ量の違いを補正しながら、学習データ２に対するＨＭＭを学習する。

【００４５】２３はクラスタ重み付き学習手段２２によって学習したＨＭＭのパラメータ（学習データ２に対応する音声パターンモデルのパラメータ）を含むＨＭＭパラメータ情報、２４はクラスタ重み付き学習手段２２からのＨＭＭパラメータ情報２３に対応するＨＭＭを格納するクラスタ重み付きモデルメモリである。また、この実施の形態２では学習するＨＭＭは、上記実施の形態１と同様に連続分布型のＨＭＭとし、１個のＨＭＭで１個の音素をモデル化するものとする。さらに、ＨＭＭのトポロジーは、図６に示すように４状態のｌｅｆｔ－ｔｏ－ｒｉｇｈｔ型とする。なお、図１と同一構成要素には同一符号を付して重複する説明を省略する。

【００４６】次に動作について説明する。この実施の形態２による音声パターンモデル学習装置の学習手順は、（１）読み上げ音声モデルの学習（発話様式別音声パターンモデル学習ステップ）、（２）自由発話音声モデルの学習（発話様式別音声パターンモデル学習ステップ）、（３）学習用トークンのクラスタリング（クラスタリングステップ）、（４）クラスタ重み係数の計算（クラスタ重み計算ステップ）、（５）クラスタ重み付きモデルの学習（クラスタ重み付き学習ステップ）の５段階に分かれる。

【００４７】先ず、読み上げ音声モデルの学習手順を説明する。読み上げ音声モデルの学習を行う際は、学習データメモリ１の出力端子Ａを学習手段３の入力端子Ａ１に接続する。また、学習手段３の出力端子Ｂを読み上げ音声モデルメモリ１３の入力端子Ｂ１に接続する。この接続状態で読み上げ音声モデルの学習を行う。

【００４８】読み上げ音声モデル学習手順１　学習手段３は、上記接続経路を介して学習データメモリ１が保持する音素テーブルを読み込み、音素テーブルに記述された順序に従って学習対象とする音素を選択する。音素テーブルは、例えば図８のように記述されている場合、先頭の音素である／ａ／を学習対象として選択する。

【００４９】読み上げ音声モデル学習手順２　学習手段３は、上述のようにして選択した音素と一致する音素表記を持つトークンの特徴ベクトルの時系列である学習データ２を学習データメモリ１から読み込む。このとき、読み上げ音声データ１－１のトークンのみ読み込む。次に、読み込んだ各トークンを用いて従来技術と同様にして、フォワード・バックワードアルゴリズムに基づいて遷移回数期待値$\gamma(k)$（ｉ，ｊ，ｔ）を計算し、（１）～（３）式によってＨＭＭのパラメータである遷移確率a_{ij}、平均値m_{ij}及び分散v_{ij}を学習する。

【００５０】学習を終了すると、学習手段３は学習を終了したモデルのパラメータである、遷移確率a_{ij}、平均値m_{ij}、分散v_{ij}及びその音素表記からなるＨＭＭパラメータ情報１２を、読み上げ音声モデルメモリ１３に送出する。読み上げ音声モデルメモリ１３では、学習を終了した上記ＨＭＭのパラメータとその音素表記とに対応するＨＭＭを保持する。

【００５１】読み上げ音声モデル学習手順３　学習手段３は、学習データメモリ１が保持する音素テーブルに記載された全ての音素に対してモデルの学習が終了するまで学習対象とする音素を音素テーブルに記述されている順序に従って選択し、上記読み上げ音声モデル学習手順２を繰り返す。このようにして読み上げ音声モデルの学習が完了すると、音素テーブルに存在する全ての音素に対しての読み上げ音声の音響的特徴をモデル化したＨＭ

Mが、読み上げ音声モデルメモリ１３に格納されることになる。

【００５２】次に、自由発話音声データモデルの学習手順を説明する。自由発話音声モデルの学習を行う際は、学習データメモリ１の出力端子Ａを学習手段３の入力端子Ａ１に接続する。また、学習手段３の出力端子Ｂを自由発話音声モデルメモリ１５の入力端子Ｂ２に接続する。この接続状態で自由発話音声モデルの学習を行う。この自由発話音声データモデルの学習手順は、上述した読み上げ音声モデルの学習手順において、読み上げ音声データ１－１のトークンを学習データ２として用いる代わりに自由発話音声データ１－２のトークンを用いてＨＭＭを学習し、学習結果を読み上げ音声モデルメモリ１３の代わりに自由発話音声モデルメモリ１５に送出して格納することによってなされる。この結果、音素テーブルに存在する全ての音素に対しての自由発話音声の音響的特徴をモデル化したＨＭＭが自由発話音声モデルメモリ１５に格納されることになる。

【００５３】次に、学習用トークンのクラスタリング手順について説明する。クラスタリングは学習データメモリ１の各トークンが読み上げ音声と自由発話音声データとのいずれに近いかをクラス分けするために行うもので、学習用トークン（学習データ２）のクラスタリングを行う際は、学習データメモリ１の出力端子Ａをクラスタリング手段１６の入力端子Ａ２に接続する。具体的なクラスタリング手順を以下に示す。

【００５４】クラスタリング手順１　クラスタリング手段１６が、上記接続経路を介して学習データメモリ１が保持する音素テーブルを読み込み、音素テーブルに記述された順序に従って、クラスタリング対象とする音素を選択する。音素テーブルが図８のように記述されている場合、先頭の音素である／ａ／をクラスタリング対象として選択する。

【００５５】クラスタリング手順２　クラスタリング手段１６は、上述のようにして選択された音素をモデル化するＨＭＭを、読み上げ音声モデルメモリ１３及び自由発話音声モデルメモリ１５の各々から読み込む。ここで、読み上げ音声モデルメモリ１３から読み込んだＨＭＭをＨＭＭＲ、自由発話音声モデルメモリ１５から読み込んだＨＭＭをＨＭＭＳと記すことにする。

【００５６】クラスタリング手順３　次に、クラスタリング手段１６は、上述のようにして選択した音素と一致する音素表記を持つ全てのトークンの特徴ベクトルの時系列とトークン番号とからなる学習データ２を学習データメモリ１から読み込む。このとき、読み上げ音声データ１－１のトークンと自由発話音声データ学習データ１－２のトークンを両方読み込む。

【００５７】クラスタリング手順４　クラスタリング手段１６は、クラスタリング手順３で読み込んだトークン番号の小さいトークンから順番にクラスタリング手順２で読み込んだＨＭＭであるＨＭＭＲとＨＭＭＳとの各々に対する尤度計算を行う。この尤度計算には、例えばトレリス又はビタビアルゴリズムを用いる。また、読み上げ音声モデルメモリ１３から読み込んだＨＭＭであるＨＭＭＲに対する尤度をＬＲ、自由発話音声モデルメモリ１５から読み込んだＨＭＭであるＨＭＭＳに対する尤度をＬＳとしたとき、ＬＲ≧ＬＳであれば、当該トークンは読み上げ音声に属することを意味する記号Ｒを当該トークン番号とともにクラスタリング結果メモリ１８に送出する。逆に、ＬＲ＜ＬＳであれば、当該トークンは自由発話音声に属することを意味する記号Ｓを当該トークン番号とともにクラスタリング結果メモリ１８に送出する。クラスタリング手段１６は、読み込んだ全てのトークンに対してＨＭＭＲとＨＭＭＳとの各々に対する尤度を計算してクラスタリングを行い、当該トークン番号とともにクラスタリング結果である記号Ｒ又はＳを、クラスタリング結果メモリ１８に格納する。

【００５８】クラスタリング手順５　クラスタリング手段１６は、学習データメモリ１が保持する音素テーブルを参照し、学習データメモリ１に存在する全ての音素に対して音素テーブルに記述されている順序に従って選択し、クラスタリング手順２～４を繰り返す。

【００５９】以上の操作によってクラスタリングが完了する。以上のように、クラスタリング手順１～５を行うことによって学習データメモリ１中の全てのトークンのクラスタリング結果１７がクラスタリング結果メモリ１８に保持されることになる。図４は実施の形態２による音声パターンモデル学習装置のクラスタリング結果メモリの内容を例示的に示す図である。図に示したように、読み上げ音声データ１－１中のトークンであるトークン番号１の／ａ／はクラスタリング結果がＲであり、読み上げ音声にクラスタリングされていることがわかる。また、読み上げ音声データ１－１中のトークンであるトークン番号１２０の／ａ／はクラスタリング結果がＳであり、自由発話音声にクラスタリングされている。これは、このトークンを含む発話が、読み上げ音声であっても自由発話音声に近かったことを意味している。

【００６０】上記のようにクラスタリングを行う理由は、読み上げ音声データ１－１中のトークンであっても自由発話音声に近い発話もあり、逆に自由発話音声データ１－２中のトークンであっても読み上げ音声に近い発話もある。このため、クラスタリングを行う前の読み上げ音声データ１－１中のトークンと自由発話音声データ１－２中のトークンという分類では、必ずしも読み上げ音声の音響的特徴を持つトークンと自由発話音声データの音響的特徴を持つトークンとによるクラスに正確に分類されていないからである。

【００６１】そこで、この実施の形態２のようにＨＭＭＲとＨＭＭＳの尤度に基づいて各トークンをクラスタリングすることにより、より正確に読み上げ音声の音響的

特徴を持つトークンと自由発話音声の音響的特徴を持つトークンとにクラス分類することができる。クラスの分類がより正確になることの利点は後述するように両クラスタの音響的特徴をより均等に反映したＨＭＭを学習することができることである。なお、クラスタリングの結果がＲのクラスタを読み上げ音声クラスタ、クラスタリングの結果がＳのクラスタを自由発話音声クラスタと呼ぶことにする。

【００６２】次にクラスタ重み係数の計算手順について説明する。

クラスタ重み係数の計算手順１　クラスタ重み計算手段１９は、出力端子Ａ及び入力端子Ａ２による経路を介して学習データメモリ１が保持する音素テーブルを読み込み、この音素テーブルに記述された順番に従ってクラスタ重み係数の計算対象とする音素を選択する。音素テーブルは、例えば図８のように記述されている場合、先頭の音素である／ａ／を重み係数の計算対象音素として選択する。

【数１０】

$$WR_C = \frac{CS_C}{CR_C + CS_C} \qquad (10)$$

【数１１】

$$WS_C = \frac{CR_C}{CR_C + CS_C} \qquad (11)$$

クラスタ重み計算手段１９は、クラスタ重み係数の計算対象として選択している当該音素の音素表記とともに重み係数WR_C、WS_Cをクラスタ重み計算結果メモリ２１に送出する。クラスタ重み計算結果メモリ２１は、クラスタ重み係数WR_C及びWS_Cを音素表記に対応させたデータとして保持する。

【００６４】クラスタ重み係数の計算手順３　クラスタ重み計算手段１９は、図８中の音素表記の順番、すなわち／ｉ／～／ｚ／の順番に上記クラスタ重み係数の計算手順２を繰り返し、各音素ごとに読み上げ音声クラスタに対する重み係数WR_Cと自由発話音声クラスタに対する重み係数WS_Cとを計算して、当該音素の音素表記とともに重み係数WR_C及びWS_Cを、クラスタ重み計算結果メモリ２１に送出する。クラスタ重み計算結果メモリ２１では、上記重み係数WR_C、WS_Cを音素表記に対応させたデータとして保持する。

【００６５】以上の操作によってクラスタ重み係数の計算が完了する。図５は実施の形態２による音声パターンモデル学習装置のクラスタ重み計算結果メモリに保持された内容を示す図である。図に示すように、音素／ａ／に対する重み係数値はWR_C＝０．２、WS_C＝０．８で、上記実施の形態１の図２における音素／ａ／に対する重み係数値ＷＲ＝０．２５、ＷＳ＝０．７５と比較するとWR_C＜ＷＲとなっており、係数値が異なっている。これは、自由発話音声データ１－２のトークンの幾つかが読み上げクラスタにクラスタリングされ、読み上げクラスタに属するトークン数が増加したことを示している。

【００６３】クラスタ重み係数の計算手順２　クラスタ重み計算手段１９は、クラスタ重み係数の計算対象として選択した音素のトークン（学習データ２）のクラスタリング結果１７をクラスタリング結果メモリ１８から読み込み、読み上げ音声クラスタに属する（即ち、クラスタリングの結果として記号Ｒが付与された）トークンの数CR_Cと自由発話音声クラスタに属する（即ち、クラスタリングの結果として記号Ｓが付与された）トークンの数CS_Cとを数える。そしてCR_CとCS_Cとを用いて、読み上げ音声クラスタに対する重み係数WR_Cと自由発話音声クラスタに対する重み係数WS_Cとを、下記（１０）、（１１）式に従って計算する。（１０）、（１１）式からわかるとおり、読み上げ音声クラスタに対する重み係数WR_Cと自由発話音声クラスタに対する重み係数WS_Cの比は、読み上げ音声クラスタに属するトークンの数CR_Cと自由発話音声クラスタに属するトークンの数CS_Cとの比に反比例する値となっている。

【００６６】次に、クラスタ重み付きモデルの学習手順を説明する。クラスタ重み付きモデルの学習を行う際は、学習データメモリ１の出力端子Ａをクラスタ重み付き学習手段２２の入力端子Ａ３に接続する。この接続状態でクラスタ重み付きモデルの学習を以下の手順で行う。

【００６７】クラスタ重み付きモデル学習手順１　クラスタ重み付き学習手段２２は、学習データメモリ１が保持する音素テーブルを読み込み、音素テーブルに記述にしたがって、学習対象とする音素を選択する。音素テーブルは、図８のように記述されている場合、先頭の音素である／ａ／を学習対象として選択する。

【００６８】クラスタ重み付きモデル学習手順２　クラスタ重み付き学習手段２２は、上記接続経路を介して選択した音素と一致する音素表記を持つ全てのトークンの特徴ベクトルの時系列とトークン番号とを学習データ２として学習データメモリ１から読み込む。このとき、読み上げ音声データ１－１のトークンと自由発話音声データ１－２のトークンを両方読み込む。ここで、読み込んだトークンの総数をＫとする。また、クラスタ重み付き学習手段２２は、クラスタ重み計算結果メモリ２１からクラスタ重み付きモデル学習手順１で選択した音素に対

する読み上げ音声クラスタの重み係数WR_Cと自由発話音声クラスタの重み係数WS_Cとを読み込む。さらに、クラスタ重み付き学習手段２２は、上述のようにして選択した音素と一致する音素表記を持つ全てのトークンのクラスタリング結果１７をクラスタリング結果メモリ１８から読み込む。

【００６９】クラスタ重み付きモデル学習手順３　次に、クラスタ重み付き学習手段２２は、読み込んだ各トークン$X(k)_1$，$X(k)_2$，$X(k)_3$，・・・，$X(k)_{T(k)}$に対する遷移回数期待値$\gamma(k)(i,j,t)$を計算する。ここで、肩の添字（ｋ）は読み込んだ全てのトークン中でｋ（ｋ＝１，２，３，・・・，Ｋ）番目のトークンであることを意味する。Ｘはトークンを構成する特徴ベクトル、Ｔ（ｋ）はｋ番目のトークンを構成する特徴ベクトルの数とする。また（ｉ，ｊ，ｔ）はトークンのｔ番目の特徴ベクトル$X(k)_t$を出力して、ＨＭＭの状態ｉから状態ｊへ遷移したことを示すものとする。この遷移回数期待値$\gamma(k)(i,j,t)$は、実施の形態１と同様に例えばフォワード・バックワードアルゴリズムを用いて計算することができる。このあと、遷移回数期待値$\gamma(k)(i,j,t)$を用いて、クラスタ重み付き遷移回数期待値$\gamma(k)_{WC}(i,j,t)$を下記（１２）式にしたがって計算する。

【数１２】

$$\gamma_{WC}^{(k)}(i,j,t) = \begin{cases} WR_C * \gamma^{(k)}(i,j,t) & (k\text{番目のトークンが読み上げ音声クラスタに属するとき}) \\ WS_C * \gamma^{(k)}(i,j,t) & (k\text{番目のトークンが自由発話音声クラスタに属するとき}) \end{cases} \quad (12)$$

このクラスタ重み付き遷移回数期待値$\gamma(k)_{WC}(i,j,t)$を用いて（１３）～（１５）式によってＨＭＭのパラメータを学習する。但し、ＨＭＭは、図６に示すような４状態であるので、ｉ＝（１，２，３）、ｊ＝（ｉ，ｉ＋１）となる。

【数１３】

$$a_{ij}^{WC} = \frac{\sum_{k=1}^{K}\sum_{t=1}^{T(k)}\gamma_{WC}^{(k)}(i,j,k)}{\sum_{k=1}^{K}\sum_{j=i}^{i+1}\sum_{t=1}^{T(k)}\gamma_{WC}^{(k)}(i,j,k)} \quad (13)$$

【数１４】

$$m_{ij}^{WC} = \frac{\sum_{k=1}^{K}\sum_{t=1}^{T(k)}\gamma_{WC}^{(k)}(i,j,t) * X_t^{(k)}}{\sum_{k=1}^{K}\sum_{t=1}^{T(k)}\gamma_{WC}^{(k)}(i,j,t)} \quad (14)$$

【数１５】

$$v_{ij}^{WC} = \frac{\sum_{k=1}^{K}\sum_{t=1}^{T(k)}\gamma_{WC}^{(k)}(i,j,k) * (m_{ij}^{WC} - X_t^{(k)})^2}{\sum_{k=1}^{K}\sum_{t=1}^{T(k)}\gamma_{WC}^{(k)}(i,j,k)} \quad (15)$$

学習を終了すると、クラスタ重み付き学習手段２２は、学習を終了したモデルのパラメータである遷移確率$a_{ij}{}^{WC}$、平均値$m_{ij}{}^{WC}$、分散$v_{ij}{}^{WC}$、及びその音素表記からなるＨＭＭパラメータ情報２３をクラスタ重み付きモデルメモリ２４に送出する。クラスタ重み付きモデルメモリ２４は、学習を終了したＨＭＭパラメータ情報２３に対応するＨＭＭを保持する。

【００７０】クラスタ重み付きモデル学習手順４　クラスタ重み付き学習手段２２は、学習データメモリ１が保持する音素テーブルに存在する全ての音素に対してモデルの学習が終了するまで学習対象とする音素を、音素テーブルに記述されている順序にしたがって選択し、上記クラスタ重み付き手順２～３を繰り返す。

【００７１】ここで、この実施の形態２による音声パターンモデル学習装置と上記実施の形態１による音声パターンモデル学習装置との違いについて説明する。上記実施の形態１では、読み上げ音声クラスタと自由発話音声クラスタの分類を読み上げ音声データ１－１のトークンか自由発話音声データ１－２のトークンかで単純に分類していた。これに対して、この実施の形態２による音声パターンモデル学習装置では、読み上げ音声モデルと自由発話音声モデルとの尤度の比較に基づいて各トークンをクラスタリングする。これにより、読み上げ音声の音響的特徴を持つトークンと自由発話音声データの音響的特徴を持つトークンとのクラス分類がより正確になる。従って、両クラスタの音響的特徴をより均等に反映したＨＭＭを学習することができる。

【００７２】以上のように、この実施の形態２によれ

ば、異なる発話様式（読み上げ音声、自由発話音声）で入力された音声データの音響的特徴を表す読み上げ音声データ１－１、自由発話音声データ１－２から抽出した学習データ２の発話様式ごとに対応する音声パターンモデルのパラメータを求め、これら発話様式ごとの音声パターンモデルのパラメータに対応する発話様式別の音声パターンモデルを用いて学習データ２をクラスタリングして各学習データ２が属する発話様式のクラスタを決定し、クラスタリングされた各発話様式のクラスタに属する学習データのデータ量の逆数$1/CR_C$、$1/CS_C$に、これらの総和が１となるように正規化したもの（各逆数$1/CR_C$、$1/CS_C$に$CR_C \cdot CS_C/(CR_C+CS_C)$を乗算する、（１０）、（１１）式参照）を、学習データ２に対する発話様式のクラスタごとのデータ量に応じたクラスタ重み係数WR_C、WS_Cとして算出し、これらクラスタ重み係数WR_C、WS_Cを用いて学習データ２の各発話様式のクラスタ間におけるデータ量の違いを補正しながら、学習データ２に対応する音声パターンモデルのパラメータを求めるので、読み上げ音声モデルと自由発話音声モデルとの尤度の比較に基づいて各学習データ２をクラスタリングすることから、上記実施の形態１による構成と比較して読み上げ音声の音響的特徴を持つ学習データ２と自由発話音声データの音響的特徴を持つ学習データ２とのクラス分類がより正確になる。これにより、両クラスタの音響的特徴を均等に反映した音声パターンモデルのパラメータを得ることができるという効果が得られる。

【００７３】また、この実施の形態２によれば、音声パターンモデルが隠れマルコフモデルであり、学習データ２から算出した遷移回数期待値$\gamma^{(k)}(i,j,t)$にクラスタ重み係数WR_C、WS_Cを乗じて、学習データ２の各発話様式のクラスタ間におけるデータ量の違いを補正したクラスタ重み付き遷移回数期待値$\gamma_{wc}^{(k)}(i,j,t)$とし、このクラスタ重み付き遷移回数期待値$\gamma_{wc}^{(k)}(i,j,t)$を用いてＨＭＭのパラメータ遷移確率a_{ij}^{wc}、平均値m_{ij}^{wc}、分散v_{ij}^{wc}を求めるので、読み上げ音声の音響的特徴を持つ学習データ２と自由発話音声データの音響的特徴を持つ学習データ２とのクラス分類精度を向上させることができる。これにより、両クラスタの音響的特徴を均等に反映した音声パターンモデルのパラメータを得ることができるという効果が得られる。

【００７４】なお、上記実施の形態２では、ＨＭＭで音素をモデル化する場合を説明したが、音節など他の音響単位でも構わない。さらに、上記実施の形態１，２では、音声パターンモデルとして、隠れマルコフモデルを使用する例を示したが、本願発明はこれに限らず、他の音声パターンモデルに適用することも可能である。

【００７５】また、上記実施の形態では、異なる発話様式として読み上げ音声と自由発話音声とについて示したが、本願発明はこれ以外の発話様式による学習データも扱うことができる。このときも、本願発明の基本概念の１つである、発話様式や発話様式のクラスタごとにおける各データ量の逆数に、これらの総和が１となるように正規化したものを重み係数、クラスタ重み係数として算出すればよい。具体的には、例えば３つの異なる発話様式の学習データに対して音声パターンモデルの学習を行う場合、発話様式ごとの学習データ量をＣ１，Ｃ２，Ｃ３とし、これらに対応する重み係数をそれぞれＷ１，Ｗ２，Ｗ３とすると、発話様式ごとの学習データ量の逆数を１／Ｃ１，１／Ｃ２，１／Ｃ３に、これらの総和が１となるように、$C1C2C3/(C1C2+C2C3+C1C3)$を乗算したものがＷ１，Ｗ２，Ｗ３となる。

【００７６】

【発明の効果】以上のように、この発明によれば、異なる発話様式で入力された音声データの音響的特徴を表す複数種類の学習データにおける発話様式ごとの各データ量の逆数にこれらの総和が１となるように正規化したものを、学習データに対する発話様式ごとのデータ量に応じた重み係数として算出し、この重み係数を用いて学習データの各発話様式間におけるデータ量の違いを補正しながら、学習データに対応する音声パターンモデルのパラメータを求めるので、異なる発話様式に属する学習データのデータ量の不均衡を補正し、各発話様式の音響的特徴を均等に反映した音声パターンモデルを学習することができ、各発話様式に対してロバストな音声パターンモデルを得ることができるという効果が得られる。

【００７７】また、文献２による従来技術と異なり、異なる発話様式で別々に音声パターンモデルの学習をすることがないので、音声パターンモデルの数を増加させない。これにより、文献２による従来技術と比較して高性能なハードウェア資源を要することがなく、コスト的にも有利である。

【００７８】この発明によれば、音声パターンモデルが隠れマルコフモデルであり、重み係数を学習データから算出した遷移回数期待値に乗じて、学習データの各発話様式間におけるデータ量の違いを補正した重み付き遷移回数期待値とし、この重み付き遷移回数期待値を用いて隠れマルコフモデルのパラメータを求めるので、異なる発話様式に属する学習データのデータ量の不均衡を補正し、各発話様式の音響的特徴を均等に反映した隠れマルコフモデルを学習することができ、読み上げ音声と自由発話音声との両方に対してロバストな隠れマルコフモデルを得ることができるという効果がある。

【００７９】この発明によれば、異なる発話様式で入力された音声データの音響的特徴を表す複数種類の学習データの発話様式ごとに対応する音声パターンモデルのパラメータを求め、これら発話様式ごとの音声パターンモデルのパラメータに対応する発話様式別音声パターンモデルを用いて学習データをクラスタリングして、各学習

データが属する発話様式のクラスタを決定し、クラスタリングした各発話様式のクラスタに属する学習データのデータ量の逆数に、これらの総和が１となるように正規化したものを、学習データに対する発話様式のクラスタごとのデータ量に応じたクラスタ重み係数として算出し、これらクラスタ重み係数を用いて学習データの各発話様式のクラスタ間におけるデータ量の違いを補正しながら、学習データに対応する音声パターンモデルのパラメータを求めるので、各発話様式別音声パターンモデルの尤度の比較に基づいて各学習データをクラスタリングすることから、上記段落００７６による構成と比較して、学習データの各発話様式の音響的特徴に対するクラス分類をより正確に行うことができ、各発話様式に対応するクラスタの音響的特徴を均等に反映した音声パターンモデルのパラメータを得ることができるという効果がある。

【００８０】この発明によれば、音声パターンモデルが隠れマルコフモデルであり、クラスタ重み係数を学習データから算出した遷移回数期待値に乗じて、学習データの各発話様式のクラスタ間におけるデータ量の違いを補正したクラスタ重み付き遷移回数期待値とし、このクラスタ重み付き遷移回数期待値を用いて隠れマルコフモデルのパラメータを求めるので、学習データの各発話様式の音響的特徴に対するクラス分類精度を向上させることができ、各発話様式に対応するクラスタの音響的特徴を均等に反映した音声パターンモデルのパラメータを得ることができるという効果がある。

【図面の簡単な説明】

【図１】　この発明の実施の形態１による音声パターンモデル学習装置の構成を示すブロック図である。

【図２】　実施の形態１による音声パターンモデル学習装置における重み計算結果メモリに保持された重み係数を示す図である。

【図３】　この発明の実施の形態２による音声パターンモデル学習装置の構成を示すブロック図である。

【図４】　実施の形態２による音声パターンモデル学習装置のクラスタリング結果メモリの内容を例示的に示す図である。

【図５】　実施の形態２による音声パターンモデル学習装置のクラスタ重み計算結果メモリに保持された内容を示す図である。

【図６】　ＨＭＭのトロポジを示す図である。

【図７】　従来の音声パターンモデル学習装置の構成を示すブロック図である。

【図８】　音素テーブルの一例を示す図である。

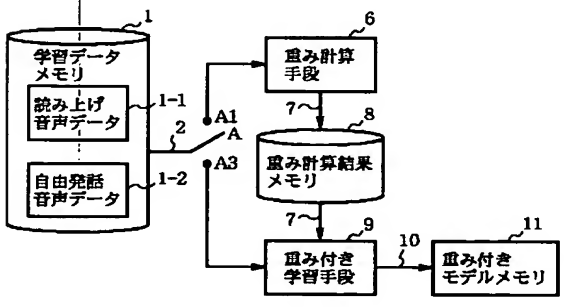
【図９】　読み上げ音声トークンテーブルを示す図である。

【図１０】　自由発話音声データトークンテーブルを示す図である。

【符号の説明】

１　学習データメモリ（学習データ記憶手段）、１－１　読み上げ音声データ（学習データ）、１－２　自由発話音声データ（学習データ）、２　学習データ、３　学習手段（発話様式別音声パターンモデル学習手段）、６　重み計算手段、７　重み係数、８　重み計算結果メモリ、９　重み付き学習手段、１０　ＨＭＭのパラメータ（音声パターンモデルのパラメータ）、１１　重み付きモデルメモリ、１２　ＨＭＭパラメータ情報（発話様式別音声パターンモデルのパラメータ）、１３　読み上げ音声モデルメモリ、１４　ＨＭＭパラメータ情報（発話様式別音声パターンモデルのパラメータ）、１５　自由発話音声モデルメモリ、１６　クラスタリング手段、１７　クラスタリング結果、１８　クラスタリング結果メモリ、１９　クラスタ重み計算手段、２０　クラスタ重み計算結果、２１クラスタ重み計算結果メモリ、２２　クラスタ重み付き学習手段、２３　ＨＭＭパラメータ情報（音声パターンモデルのパラメータ）、２４　クラスタ重み付きモデルメモリ。

【図１】



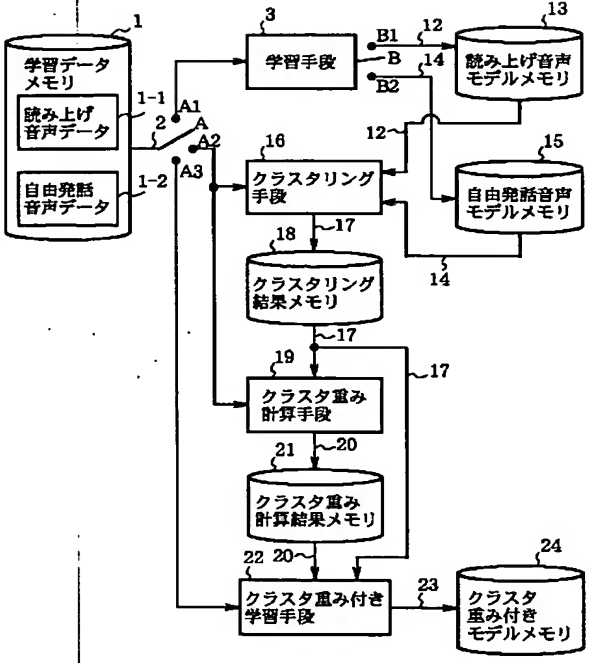
【図２】

番号	音素名	重みWR（読み上げ音声データ）	重みWS（自由発話音声データ）
1	a	0.25	0.75
2	i	0.5	0.5
3	u	0.4	0.6
…	…	…	…
25	z	0.7	0.3

【図８】

番号	音素名
1	a
2	i
3	u
…	…
25	z

【図3】



【図4】

トークン番号	音素表記	クラスタリング結果
1	a	R
2	a	R
…	…	…
120	a	S
121	i	S
122	i	R
…	…	…
300	i	R
…	…	…
915	z	R
916	z	S
…	…	…
1020	z	S
1021	a	S
1022	a	S
…	…	…
1060	a	R
1061	i	S
1062	i	S
…	…	…
1100	i	S
…	…	…
1521	z	R
1522	z	S
…	…	…
1580	z	S

【図9】

トークン番号	音素表記
1	a
2	a
…	…
120	a
121	i
122	i
…	…
300	i
…	…
915	z
916	z
…	…
1020	z

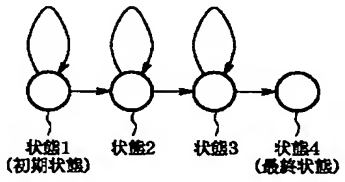
【図10】

トークン番号	音素表記
1021	a
1022	a
…	…
1060	a
1061	i
1062	i
…	…
1100	i
…	…
1521	z
1522	z
…	…
1580	z

【図5】

番号	音素名	重みWRc (読み上げクラスタ)	重みWSc (自由発話クラスタ)
1	a	0.2	0.8
2	i	0.5	0.5
3	u	0.4	0.6
…	…	…	…
25	z	0.7	0.3

【図6】



【図7】